A Comparison of Estimators for a Two-parameter Hyperbola

By D. COLQUHOUN

University College London

SUMMARY

The hyperbola $Y = Vx/(K+x)$ occurs frequently in biochemistry (the Michaelis–Menten equation) and related subjects. The estimates of the parameters V and K obtained by the method of least squares (applied to both the observations and their logarithms) are compared with the conventional estimates from linear transformations by using simulated normally and lognormally distributed observations.

With homoscedastic normal observations the least-squares (and thus maximum-likelihood) estimates of V and K have substantially smaller variance and less bias than the estimates from the best of the linear transformations unless the experimental results are very precise.

When the coefficient of variation of the observations is constant none of the methods tested stands out as uniformly the best. There is a strong positive correlation between the estimates of V and K whichever method of estimation is used.

1. THE PROBLEM

THE problem to be considered is how best to estimate the parameters, V and K, of the rectangular hyperbola

$$Y = \frac{Vx}{K+x}, \tag{1}$$

where Y is the dependent variable and x the independent variable. This equation occurs often in biological sciences; for example in enzymology Y is the initial velocity of an enzyme-catalysed reaction, x is substrate concentration, V is maximum velocity (when $x \to \infty$) and K is the Michaelis constant.

Equation (1) can be transformed to produce linear plots in three ways:

$$Y = V - K\left(\frac{Y}{x}\right), \tag{2}$$

$$\frac{x}{Y} = \frac{x}{V} + \frac{K}{V}, \tag{3}$$

$$\frac{1}{Y} = \frac{1}{V} + \frac{K}{V}\left(\frac{1}{x}\right). \tag{4}$$

In practice one of these linear plots, usually $1/Y$ against $1/x$ according to equation (4), is fitted by eye or unweighted least squares to obtain estimates of V and K from the slope and intercept of the fitted line. Dowd and Riggs (1965) have compared the estimates of the parameters obtained by these three transformations using simulation methods. They found that the most commonly used transformation (equation (4), which is known to biochemists as the Lineweaver–Burk plot) is outstandingly the worst of the three methods.

Several improvements on the linear transformations have been proposed (see, for example, Wilkinson, 1961; Bliss and James, 1966; Cleland, 1963, 1967). It has been assumed in almost all discussions of the problem that the best (unbiased and minimum variance in small samples) estimates will be given by the method of least squares, i.e. by minimizing $\sum w(Y-Y^*)^2$ where Y^* is calculated from equation (1), Y is an observed value of the dependent variable and w is the reciprocal of the variance of Y. There is, however, no evidence that this is so even when the variance of Y is constant and the error in Y is normally distributed. Also the behaviour of the least-squares estimates when the variance of Y is not constant, or Y not normally distributed, is of considerable interest in selecting a method of estimation, since in practice it is often not known whether the variance is constant and the distribution of the observations is virtually never known. Furthermore, even when the least-squares estimates are, in some sense, the best estimates, none of the discussions referred to has considered whether or not the improvement on the linear transformations is sufficiently large to be of practical importance. The study of Dowd and Riggs is therefore extended, in this paper, to compare also the method of least squares, applied to both the observations and their logarithms using normally and lognormally distributed simulated observations.

In the case of normal homoscedastic observations the least-squares estimates are also the maximum-likelihood estimates (see, for example, Draper and Smith, 1966, p. 265).

2. Methods

2.1. *The Model for Simulation*

Dowd and Riggs (1965) chose an experimental arrangement representative of current practice by inspection of 28 sets of results in six consecutive issues of the *Journal of Biological Chemistry*. The arrangement chosen by them consisted of five observations of Y, one at each of the values $x = 2{\cdot}5$, $5{\cdot}0$, $10{\cdot}0$, $20{\cdot}0$ and $40{\cdot}0$, with $V = 30$ and $K = 15$. The same design has been used throughout this work.

A digital computer was used to generate random normally distributed values of Y with expectations $30x/(15+x)$, and with various specified standard deviations. Five such values constituted one simulated experiment and estimates of V and K were obtained from it by each of the methods under investigation. Normally distributed random variables were generated using a procedure based on the algorithms of Pike and Hill (1965) and Pike (1965).

2.2. *Estimation Methods*

Straight lines were fitted by unweighted least squares for each of the linear transformations. Least-squares estimates were obtained by finding the values of V and K (constrained to be positive) needed to minimize

$$\sum (Y-Y^*)^2. \tag{5}$$

Least-squares estimates using the logarithmic transformation were obtained by minimizing

$$\sum (\log Y - \log Y^*)^2. \tag{6}$$

The minimization was carried out by a direct search method, *patternsearch*, written by M. Bell of the University of London Institute of Computer Science on the basis of the work of Hooke and Jeeves (1961). The procedure given by Bell and Pike (1966) is very similar to patternsearch. Bell's patternsearch procedure, and an example of the contours for the sum of squares surface (eq. 5), are included in "Lectures on statistics with applications in biology and medicine" (Colquhoun, D. In preparation). Patternsearch has been successfully used for the fitting of more complex curves with five or six parameters (Colquhoun, 1968). The initial guesses for the minimization were (30, 15) for (V, K) in most runs but, as long as V and K were constrained to be positive (as they must be on physical grounds), it was found that initial guesses of (1, 1), (1, 300), (300, 1) and (300, 300) all resulted in attainment of the same minimum as (30, 15) in a series of 20 simulated experiments chosen to give a very wide range of estimates. The minimum was virtually always reached after fewer than 220 evaluations of equation (1).

Generality of results

The results given are independent of the particular values (30 and 15) chosen for the parameters. If V is multiplied by any factor then as long as the standard deviation of Y is altered by the same factor all the estimates of V are simply changed by this factor. Similarly if K and the values of x used are multiplied by any factor, the estimates of K are altered by the same factor. The results will, however, only apply to experiments in which the values of x chosen are the same in relation to K, and the standard deviations the same in relation to V, as in the model used for simulation.

3. Results

To facilitate comparison the results are presented in a form similar to that used by Dowd and Riggs (1965).

The mean-square error in Table 1 is calculated as the sum of the squares of deviations of individual estimates from their true values divided by the number of estimates (N).

The results obtained using the three linear transformations confirm those of Dowd and Riggs (1965) and they have been included to enable comparisons to be made with the least-squares estimates.

3.1. *Normally Distributed Observations with Small Error of Constant Magnitude*

A value of 0·2 was used for the standard deviation (σ) of Y at each value of x so that the coefficient of variation of Y decreases from 4·78 per cent at the lowest x value to 0·91 per cent at the highest.

From the distribution of the estimates of V and K shown in Fig. 1 and the results in Table 1, it can be seen that the unweighted least-squares (and maximum-likelihood) method gives the best estimates though they are only slightly better than the estimates obtained by the best of the linear transformations (in this case x/Y against x). These results hold good whether the criterion used to compare methods of estimation is their variance, bias or mean-square error.

There is a very high positive correlation between the estimates of V and K whatever method of estimation is used. Similar high correlations were also seen using the other models for error as shown in Table 1.

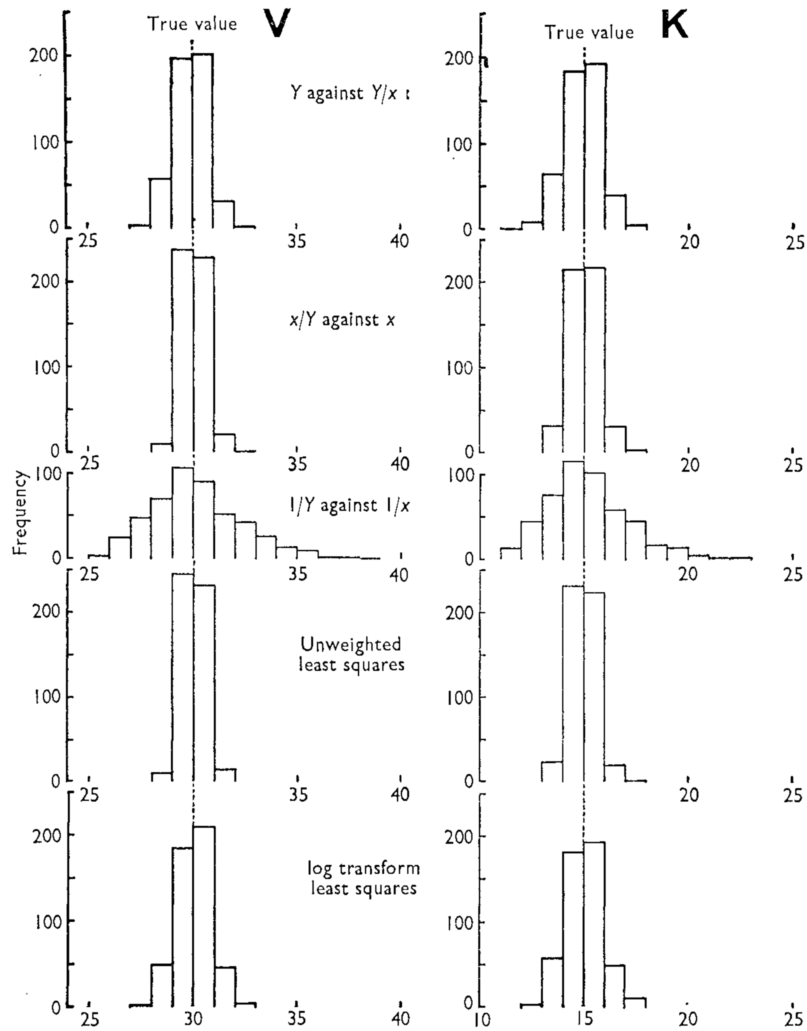

FIG. 1. Distributions of estimates of V (left-hand column) and K (right-hand column) obtained by five methods from 500 replicate simulated experiments. Model (1): Normally distributed observations with small constant variance.

TABLE 1

Mean, variance (Var) and mean-square error (MSE) of the estimates of V and K obtained in N simulated experiments, and the product moment correlation coefficient (r) between the estimates. The methods of estimation are (1) *Y against Y/x*, (2) *x/Y against x*, (3) 1/*Y against* 1/*x*, (4) *unweighted least squares*, (5) *log transform least squares*

Model	*Method*	*V*			*K*			*r*
		Mean	*Var*	*MSE*	*Mean*	*Var*	*MSE*	
1	(1)	29·9	0·60	0·61	14·9	0·78	0·79	0·97
Normal	(2)	30·0	0·30	0·30	15·0	0·44	0·44	0·93
$\sigma = 0{\cdot}2$	(3)	30·2	4·82	4·86	15·2	3·85	3·88	0·99
$N = 500$	(4)	30·0	0·26	0·26	15·0	0·35	0·35	0·93
	(5)	30·0	0·63	0·63	15·0	0·83	0·83	0·97
2	(1)	28·0	12·2	16·2	13·6	13·9	15·9	0·94
Normal	(2)	31·0	18·9	19·9	16·7	34·9	37·6	0·97
$\sigma = 1{\cdot}0$	(3)	†	†	†	†	†	†	1·0
$N = 748$‡	(4)	30·4	7·5	7·6	15·6	10·5	10·8	0·93
	(5)	31·1	32·4	33·5	16·7	48·9	51·6	0·98
3	(1)	29·7	2·89	2·96	14·8	1·91	1·95	0·93
Normal	(2)	29·9	3·99	3·99	15·0	2·88	2·88	0·95
$CV = 4{\cdot}7\%$	(3)	30·3	7·53	7·61	15·3	4·78	4·88	0·97
$N = 750$	(4)	30·0	4·58	4·58	15·0	4·10	4·10	0·95
	(5)	30·0	2·99	2·99	15·1	2·01	2·01	0·93
4	(1)	24·8	52·5	79·8	10·4	33·7	54·6	0·92
Normal	(2)	33·6	350	362	19·9	503	526	0·96
$CV = 23{\cdot}3\%$	(3)	†	†	†	†	†	†	1·0
$N = 750$	(4)	†	†	†	†	†	†	1·0
	(5)§	32·6	164	171	18·1	151	161	0·95
5	(1)	29·8	3·53	3·56	14·8	2·28	2·32	0·94
Lognormal	(2)	30·1	4·60	4·60	15·1	3·17	3·18	0·95
$CV = 4{\cdot}7\%$	(3)	30·2	7·89	7·92	15·2	5·02	5·04	0·97
$N = 500$	(4)	30·2	5·38	5·39	15·1	4·59	4·60	0·95
	(5)	30·1	3·64	3·65	15·1	2·36	2·36	0·94
6	(1)	26·0	63·2	79·4	11·2	40·8	55·0	0·94
Lognormal	(2)	33·6	381	394	18·8	299	313	0·98
$CV = 23\%$	(3)	†	†	†	†	†	†	0·99
$N = 750$	(4)	†	†	†	†	†	†	1·0
	(5)	33·5	340	352	18·2	248	257	0·98

† Values too large to be meaningful.

‡ Two experiments which gave near infinite estimates have been omitted. This has a noticeable effect on the results only for methods (2) and (5).

§ Two experiments which gave near infinite estimates using method (5) have been omitted.

3.2. *Normally Distributed Observations with Large Error of Constant Magnitude*

$\sigma = 1{\cdot}0$ so that the coefficient of variation decreases from 23·3 per cent to 4·6 per cent.

From the results shown in Fig. 2 and Table 1 it can be seen that it is in this case that the unweighted least-squares (and thus maximum-likelihood) method shows the

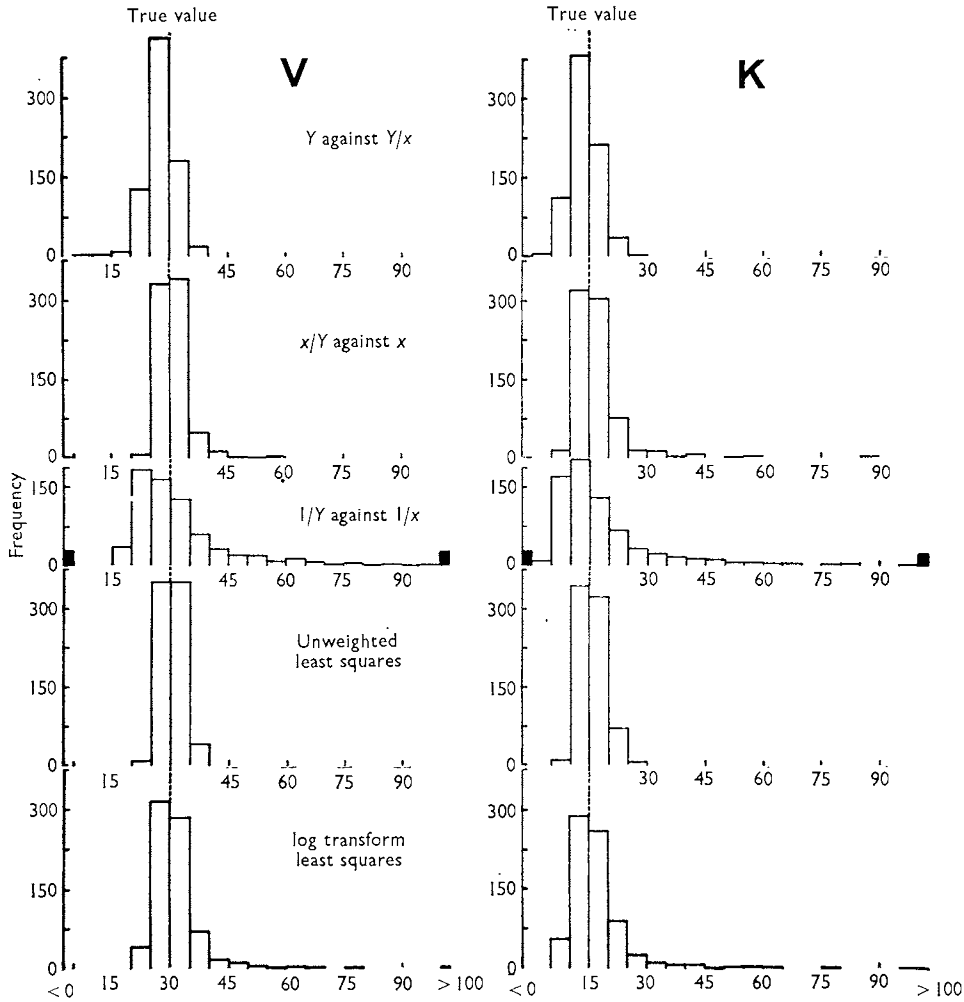

FIG. 2. Distributions of estimates of V (left-hand column) and K (right-hand column) obtained by five methods from 750 replicate simulated experiments. Model (2): Normally distributed observations with large constant variance.

biggest advantage over the other methods. As found by Dowd and Riggs (1965), the best of the linear transformations in this case is Y against Y/x, but the unweighted least-squares estimates are substantially better than those obtained by this method, having both smaller variance and less bias.

3.3. *Normally Distributed Observations with Small Error, Increasing with Y*

The standard deviation of Y was 0·2 for the lowest value of x and increased in direct proportion to the expectation of Y so that the coefficient of variation of Y remained constant at 4·7 per cent.

The results in Fig. 3 and in Table 1 show that the estimates obtained by the log transform least-squares method have less bias but slightly larger variance than the

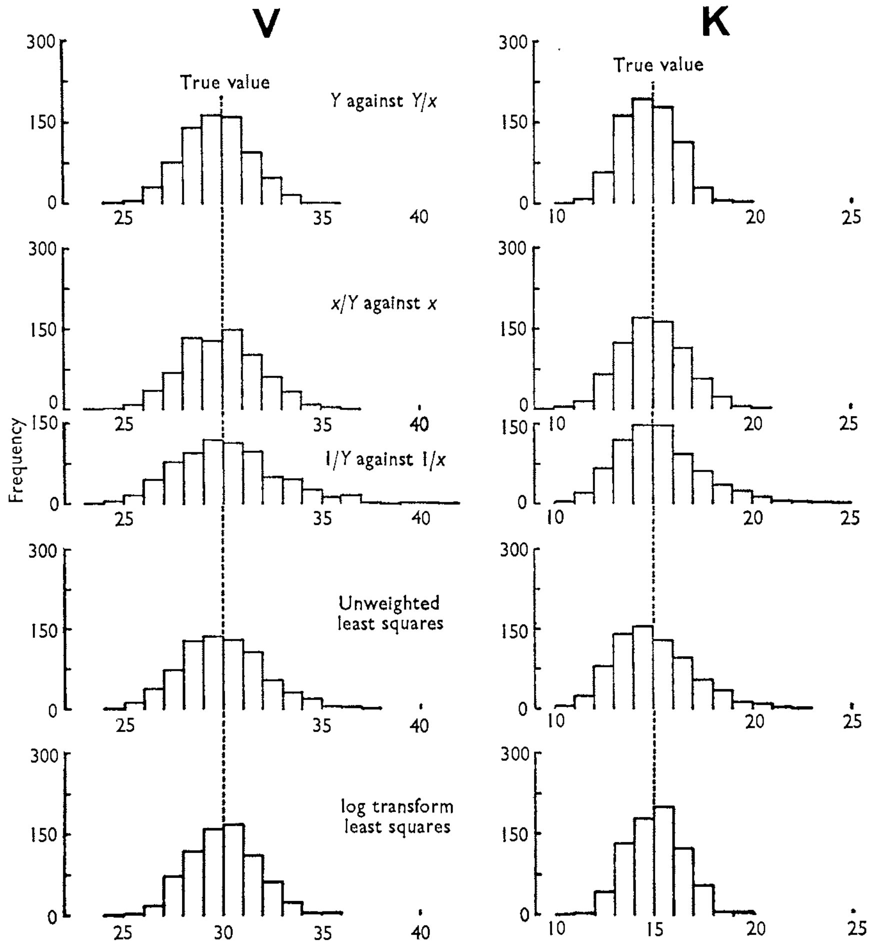

FIG. 3. Distributions of estimates of V (left-hand column) and K (right-hand column) obtained by five methods from 750 replicate simulated experiments. Model (3): Normally distributed observations with small constant coefficient of variation.

estimates found using the best linearizing transformation (Y against Y/x in this case) so the mean square errors are similar for both methods. The unweighted least-squares method, when used with observations with non-constant variance, is seen to give results which are worse than those obtained by either of the two best linear transformations. As usual the estimates obtained by the double reciprocal plot, equation (4), were worse than those by any other method.

3.4. *Normally Distributed Observations with Large Error, Increasing with Y*

At the lowest x value $\sigma = 1$, increasing so that the coefficient of variation of Y was constant at 23·3 per cent. This model represents experiments with unusually large errors.

From Fig. 4 and Table 1 it can be seen that the log transform least-squares estimates tail towards higher values, giving a positively skewed distribution with occasional

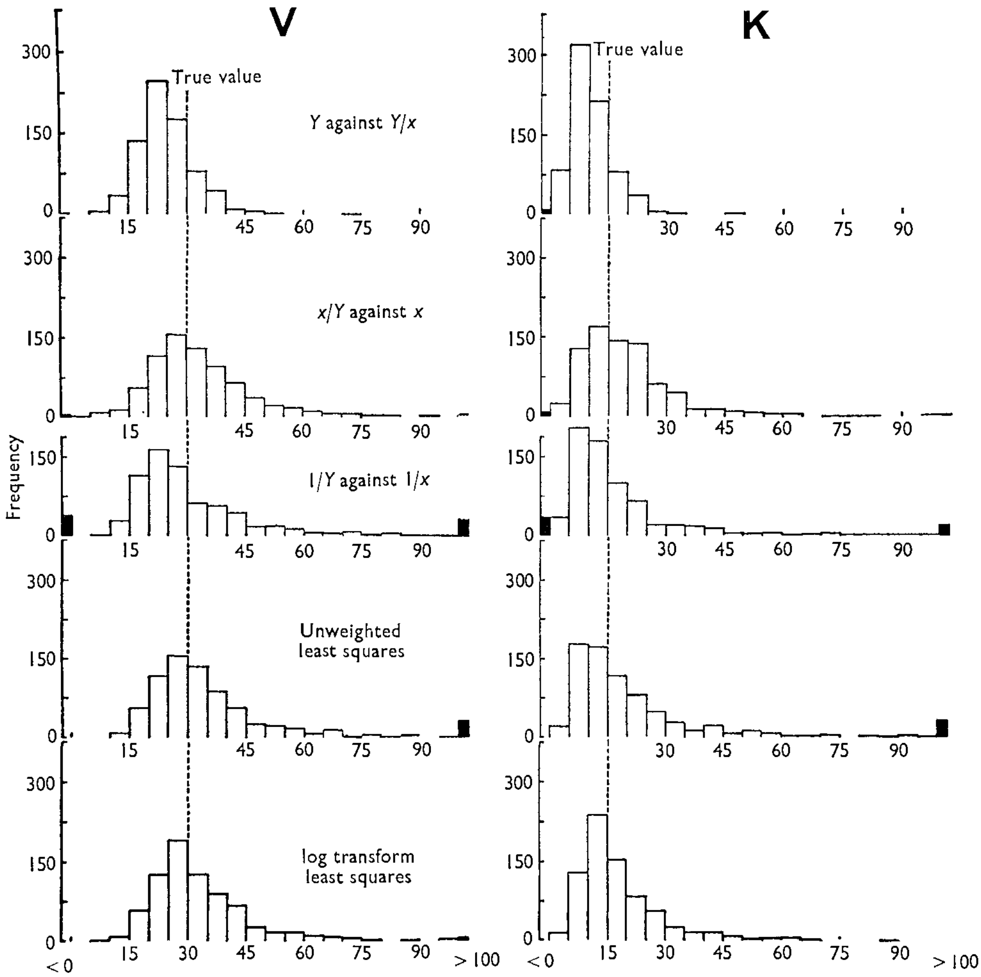

FIG. 4. Distributions of estimates of V (left-hand column) and K (right-hand column) obtained by five methods from 750 replicate simulated experiments. Model (4): Normally distributed observations with large constant coefficient of variation.

excessively high estimates. As a result there is some positive bias in the estimates. The only method which did not give occasional estimates tending towards infinite values was the plot of Y against Y/x and the estimates by this method have a smaller variance than those by any of the other methods. They are, however, very biased towards low values. Using Y against Y/x more than 80 per cent of estimates of both V and K were less than the true values, whereas using log transform least squares the figure was about 50 per cent.

3.5. *Lognormally Distributed Observations with Small Error*

Values of Y the logarithms of which are normally distributed were obtained by generating normally distributed random variables with means of $\log\{30x/(15+x)\}$, and with a constant standard deviation of 0·0467. The antilogarithms of these variables were taken as the simulated observations. The expectation of Y itself should not in this case be exactly $30x/(15+x)$ but somewhat larger. The standard deviation of Y increases in such a way that the coefficient of variation of Y is approximately constant at 4·7 per cent.

3.6. *Lognormally Distributed Observations with Large Error*

Experiments were simulated as described for model (5) except that the standard deviation of the normally distributed log Y was constant at 0·233 so that the coefficients of variation of the lognormally distributed Y were approximately constant at 23 per cent. In a typical run of 250 experiments all five coefficients of variation were between 21·3 and 24·3 per cent.

In spite of the fact that models (5) and (6) are not strictly comparable with the previous ones it can be seen from Table 1 that the results obtained using lognormally distributed observations were qualitatively similar to those obtained when the distribution of observations is normal (models (3) and (4)). The histograms of the distributions of the estimates have therefore not been given separately.

4. Discussion

The results which have been presented indicate that, in the case when the error of the observations is reasonably constant, estimation of V and K by the method of least squares gives a worthwhile improvement over the best of the linear transformations unless the experiments are very precise. Furthermore, which of the linear transformations is the best depends on whether the error is large or small, whereas the method of least squares gives the best estimates, in this case, whatever the size of the error.

There is no doubt that the double reciprocal plot of $1/Y$ against $1/x$ gives uniformly the worst estimates, as Dowd and Riggs (1965) found.

In the case where the standard deviation of the dependent variable increases in proportion to its expectation so that the coefficient of variation of the observations is constant the best method of estimation is not so obvious. When the coefficient of variation has the reasonable value of 4·6 per cent there is little to choose between the log transform least-square estimates and those obtained by the Y against Y/x transformation. The latter estimates are, as usual, the most biased towards low values but have an only very slightly smaller variance. It is of interest that in this case it would certainly be better to use a linear transformation (except for $1/Y$ against $1/x$) than to use unweighted least squares.

When the coefficient of variation has the perhaps unrealistically large value of 23·3 per cent, the distribution of estimates obtained using the log transform least-squares method shows tailing towards high values and two of the 750 experiments produced estimates of V and K which tended towards infinity. These have been omitted as noted in the footnote to Table 1. However, about 50 per cent of estimates were below the true value for both V and K. The unweighted least-squares method produced much more severe tailing, there being many very large estimates. Again only the plot of Y against Y/x could be relied upon not to produce occasional absurdly high estimates though in this case it is extremely biased and more than 80 per cent of estimates were below the true value for both V and K.

The log transform least-squares method has been considered by Nelder (1966, 1968), who describes a non-iterative approximation to it which is different from all of the methods used here.

The reason why the plots of $1/Y$ against $1/x$ and x/Y against x can produce such large errors in particular experiments is that the intercept (in the former case) and slope (in the latter) of the linear plot occur in the *denominator* of the expressions for V and K. It is quite possible for the slope or intercept occasionally to be close to zero so occasional huge (positive or negative) estimates of V and K are obtained. However when Y is plotted against Y/x the slope of the line is $-K$ and the intercept is V. Since neither the slope nor the intercept occurs in the denominator absurdly large estimates of V and K are never found using this method and in the case of model (4) this method actually gives a smaller scatter of estimates than the log transform least-squares method.

Although the plot of Y against Y/x never gives huge errors it is badly biased towards low values if the experimental errors are at all large, and this is serious from the experimental point of view. If only one experiment were performed then the mean-square error might reasonably be taken as the criterion of the "best" estimate. However, this is virtually never the case. The experiment is almost always repeated several times. This being so it is arguable that large bias will usually be a far more serious defect in an estimation method than large variance, since only the latter can be detected and reduced by repetition of the experiment.

The least-squares methods can also produce estimates of V and K that tend towards infinity. This will occur when the experimental points are best fitted by a straight line which is characterized by infinite values of V and K. This would be expected to occur most frequently by chance when the error is large and increasing with Y (models (4) and (6)) and it is in just these cases that a substantial number of absurdly large estimates were found. If larger values of x had been included in the model it would be less likely that the observations would appear to lie on a straight line so presumably absurdly large estimates would be rarer.

Whatever method of estimation is used, a high positive correlation is found between the estimates of V and K so that if the estimate of one of them is too large it is very likely that the estimate of the other will also be too large.

Acknowledgements

I am very grateful to Professor B. G. Greenberg, Mr I. D. Hill and Mr N. W. Please for their help and advice, and to Mr M. Bell for making available his patternsearch procedure.

References

Bell, M. and Pike, M. C. (1966). Remark on algorithm 178. Direct search. *Communs Ass. Comput. Mach.*, **9**, 684.

Bliss, C. I. and James, A. T. (1966). Fitting the rectangular hyperbola. *Biometrics*, **22**, 537–602.

Cleland, W. W. (1963). Computer programmes for processing enzyme kinetic data. *Nature*, **198** 463–465.

—— (1967). The statistical analysis of enzyme kinetic data. *Adv. Enzymol.*, **29**, 1–32.

COLQUHOUN, D. (1968). The rate of equilibrium in a competitive *n* drug system and the auto-inhibitory equations of enzyme kinetics. Some properties of simple models for passive sensitization. *Proc. Roy. Soc.*, **B170**, 135–154.

DOWD, J. E. and RIGGS, D. S. (1965). A comparison of estimates of Michaelis–Menten kinetic constants from various linear transformations. *J Biol. Chem.*, **240**, 863–869.

DRAPER, N. R. and SMITH, H. (1966). *Applied Regression Analysis.* New York: Wiley.

HOOKE, R. and JEEVES, T. A. (1961). "Direct search" solution of numerical and statistical problems. *J. Ass. Comput. Mach.*, **8**, 212–229.

NELDER, J. A. (1966). Inverse polynomials, a useful group of multi-factor response functions. *Biometrics*, **22**, 128–141.

NELDER, J. A. (1968). Weighted regression, quantal response data, and inverse polynomials. *Biometrics*, **24**, 979–985.

PIKE, M. C. (1965). Algorithm 267. Random normal deviate. *Communs Ass. Comput. Mach.*, **8**, 606.

PIKE, M. C. and HILL, I. D. (1965). Algorithm 266. Pseudo-random numbers. *Communs Ass. Comput. Mach.*, **8**, 605–606.

WILKINSON, G. N. (1961). Statistical estimations in enzyme kinetics. *Biochem. J.*, **80**, 324–332.